

# Comprehensive Analysis of the Expression and Prognosis for Signal Transducer and Activator of Transcription Family of Genes in Acute Myeloid Leukemia

Tong Qin (✉ [hhqintong@henu.edu.cn](mailto:hhqintong@henu.edu.cn))

Huaihe Hospital of Henan University <https://orcid.org/0000-0003-2085-599X>

Yunli Shao

First Affiliated Hospital of Henan University

Hongmian Zhao

Huaihe Hospital of Henan University

---

## Research

**Keywords:** acute myeloid leukemia, STATs, prognosis, TCGA

**Posted Date:** August 30th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-832374/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

In recent years, it has become clear that *signal transducer and activator of transcription (STAT) family of genes* play an important role in cancer. However, the expression level, genetic variation, molecular mechanism, and biological function of different STATs in acute myeloid leukemia (AML) and its relationship with prognosis and immune infiltration in patients with AML have not been fully elucidated.

## Results

We found that the expression levels of STAT1, STAT2, STAT4, and STAT6 were higher in AML than in normal tissues, whereas the expression levels of STAT3, STAT5A, and STAT5B were lower in the former than in the latter. Survival analysis revealed that high transcription level of STAT6 was associated with low overall survival (OS) in patients with AML. Conversely, high STAT5B level predicted superior OS in these patients. Cox multivariate analysis as well as least absolute shrinkage and selection operator (LASSO) regression screening showed that STAT5B expression is a good independent prognostic factor of OS. The functions of STAT5B are mainly related to the occurrence, development and microenvironment of AML. We also discovered that the expression of STAT5B was significantly correlated with immune infiltrates in AML.

## Conclusions

STAT5B may be used as a novel predictive biomarker and immunotherapy target for AML patients.

## Background

Acute myeloid leukemia (AML) is a collection of hematologic malignancies with remarkable cytogenetic abnormalities and gene mutations(1). Traditionally, cytogenetics has been approved to be the important prognostic parameter in AML, due to recent advances in next-generation sequencing, some genes frequently mutated in AML have been identified that are meaningful for clinical practice(2). Molecular genetic data are increasingly being utilized to inform disease classification and risk stratification of patients(3). For instance, *NPM1* and biallelic *CEBPA* mutations are favorable prognostic markers for patients with cytogenetically normal AML(4). Mutations of *FLT3-ITD*(5), *TP53*(6), *RUNX1*(7), *ASXL1*(8), and *DNMT3A*(9) consistently confer poor prognosis to AML patients. However, the prognosis of some patients is still complex.

Signal transducers and activators of transcription (STAT) are a family of genes encoding intracellular transcription factors that can regulate cell proliferation, differentiation, and cell cycle progression in cancer. To date, seven STATs have been identified in the human genome: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B and STAT6. They play a complex and important role in inflammation, proliferation, differentiation, cell survival and immune response(10). Among the seven family members, STAT1 is

known to be essential for the induction of Th1 cells downstream of Interferon-gamma(11). STAT1 has also been demonstrated to inhibit the differentiation of regulatory T cells and promote inflammatory diseases such as graft-versus-host disease(12). These pro-inflammatory properties of STAT1 are also vital for infection control(13). Studies have indicated that STAT1, as a tumor suppressor, can regulate apoptosis by inhibiting non-transcriptional mechanisms such as anti-apoptotic protein nuclear factor-kappa B(14). Additional research has been reported that by inducing p53 phosphorylation to regulate p53 activity, STAT1 promotes cell apoptosis(15). Comprehensive meta-analysis shows that STAT1 has prognostic value for solid cancer patients(16). STAT2 is a pivotal regulator of the interferon-regulated anti-viral response, and plays a vital role in host defenses against viral infections(17). STAT3 is one principal regulator involved in inflammation and carcinogenesis. As a transcription factor, STAT3 regulates a set of genes implicated in cancer cell survival, proliferation, angiogenesis, invasion, metastasis, drug resistance and immune evasion. Thus, a large amount of evidence strongly supports STAT3 as a cancer treatment target, such as degradation of STAT3 protein(18), block STAT3-dependent transcriptional activity(19) and reduces STAT3 phosphorylation(20). In vivo and in vitro studies in recent decades have suggested that STAT4 may induce inflammation and autoimmune diseases, inhibit tumor growth or promote tumors via regulating various facets of the innate and adaptive immune responses(21). STAT5 regulates a large number of target genes involved in cell response to cytokines and growth factors, and was selected as a target gene due to its crucial role in the signal transduction of multiple cancer cell types(22). STAT6 participates in the regulatory circuit that controls the proliferation and differentiation of T helper cells(23).

Next-generation sequencing technologies such as RNA-sequencing have become standard for querying gene expression in tissues and cells(24). Gene expression patterns in bioinformatics data sets help provide information about expression differences, relevant clinical features and functions, and clinical outcomes related to these genes. In this study, we used various large-scale bioinformatics databases to conduct deeply into and comprehensive bioinformatics exploration of the expression of STATs in AML. The expression and mutation of different STATs were analyzed to determine the expression patterns, potential functions and distinct prognostic values in AML.

## Results

### Functional enrichment analysis of STATs in human species

In this study, we used Metascape to perform pathway enrichment analysis of STATs in human species. For each STATs, pathway and process enrichment analysis has been carried out with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). All genes in the genome have served as the enrichment background. The most statistically significant term within a cluster is chosen to represent the cluster. Top 5 clusters with their representative enriched terms were Interleukin-20 family signaling, Interleukin-21 signaling, Thymic Stromal Lymphopoietin Signaling Pathway, Hepatitis B infection, and Interleukin-4 and Interleukin-13 signaling (Fig. 1A). To further capture the relationships between the terms, a subset of enriched terms have been selected and rendered as a network plot, where terms with a

similarity > 0.3 are connected by edges. We select terms with the best p-values from each of the 20 clusters, with the constraint that there are no more than 15 terms per cluster and no more than 250 terms in total. Each node of the network represents an enriched term and is colored by its cluster identification (Fig. 1B). Next, we determined the relationship between STATs and human diseases by using the ontology category DisGeNET. The first 20 enriched clusters are presented in Fig. 1C. Eight of them are related to hematopoietic disorder, such as lymphoma and myeloid leukemia.

## Expression levels of STATs in AML patients

We examined the expression of STATs in 173 cases of AML and 70 cases of normal tissues. The expression level of STATs was searched and analyzed from The Cancer Genome Atlas (TCGA) database. The results indicated that the expression levels of STAT1, STAT2, STAT4, and STAT6 were higher in AML than in normal tissues, and the expression levels of STAT3, STAT5A, and STAT5B were lower in the former than in the latter (Fig. 2). The Mann-Whitney U test (Wilcoxon rank-sum test) showed that the difference was statistically significant ( $P < 0.001$ ).

## Correlation between the expression of STATs and the survival of AML patients

We further explored the critical efficiency of STATs in the survival of patients with AML. Kaplan-Meier analysis was used to analyze 151 samples from the TCGA dataset to assess the correlation between the expression levels of STATs and the survival of AML patients. Our studies showed that AML patients with high expression of STAT5B were significantly associated with longer overall survival (OS) ( $P = 0.002$ ). In contrast, AML patients with high expression of STAT6 were significantly related with shorter OS ( $P = 0.048$ ), while STAT1-4, and 5A transcription levels were not associated with OS in AML patients (Fig. 3).

To assess the prognostic significance of clinical and molecular variables, univariate and multivariate COX regression analyses were conducted, encompassing expression levels of STAT5B (high vs. low), STAT6 (high vs. low), age ( $\geq 60$  vs.  $<60$  years), peripheral white blood cell count ( $\geq 20$  vs.  $<20 \times 10^9/L$ ), percentage of bone marrow blasts ( $\geq 20$  vs.  $<20\%$ ), percentage of peripheral blood blasts ( $\geq 70$  vs.  $<70 \times 10\%$ ), cytogenetic risk (favorable and intermediate vs. poor), *FLT3* (positive vs. negative) and *NPM1* (positive vs. negative). The results of this analysis were summarized in Table 1. Univariate analysis indicated that with high expression of STAT5B, patients had longer OS ( $P = 0.002$ ) than those who with low expression of STAT5B. Age  $\geq 60$  years had an unfavourable effect on OS ( $P < 0.001$ ). High expression of STAT6 and poor cytogenetic risk also contributes to poor OS ( $P = 0.048$ ;  $P = 0.018$ , respectively). Based on multivariate analyses, high STAT5B expression significantly associated with longer OS ( $P < 0.001$ ), while Age  $\geq 60$  years was also an independently risk factor for OS ( $P < 0.001$ ).

Table 1  
Univariate and multivariate analyses for OS

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Age (> 60 vs. <=60)	3.333 (2.164–5.134)	< 0.001	3.180 (2.015–5.019)	< 0.001
WBC count(x10 <sup>9</sup> /L) (> 20 vs. <=20)	1.161 (0.760–1.772)	0.490		
BM blasts(%) (> 20 vs. <=20)	1.165 (0.758–1.790)	0.486		
PB blasts(%) (> 70 vs. <=70)	1.230 (0.806–1.878)	0.338		
Cytogenetic risk (Favorable&Intermediate vs. Poor)	0.553 (0.338–0.905)	0.018	0.648 (0.386–1.089)	0.101
FLT3 mutation (Positive vs. Negative)	0.787 (0.496–1.248)	0.309		
NPM1 mutation (Positive vs. Negative)	0.879 (0.546–1.416)	0.596		
STAT5B (High vs. Low)	0.499 (0.324–0.769)	0.002	0.443 (0.285–0.689)	< 0.001
STAT6 (High vs. Low)	1.533 (1.004–2.341)	0.048	1.376 (0.880–2.152)	0.162
Abbreviations: OS, overall survival; CI, confidence interval; WBC, white blood cell; BM, bone marrow; PB, peripheral blood.				

The least absolute shrinkage and selection operator (LASSO) Cox regression model was used to select the most predictive gene as a prognostic indicator.  $\lambda$  was selected when the median of the sum of squared residuals was the smallest. The model is a RiskScore formula containing multiple genes. Each gene is under a weight. A negative number means that the gene is a protective gene, and a positive number means that the gene is a dangerous gene. STAT1, STAT4, STAT5A, STAT5B and STAT6 were identified as potential predictors for AML. The weights of the five genes are 0.0527, 0.2629, 0.2353, -1.0264 and 0.6362, respectively. The results indicating that STAT5B could serve as the most prognostic biomarker in prognosis of AML (Fig. 4).

## Predicted Functions and pathways of STAT5B in AML patients

In order to further evaluate the role of STAT5B in AML, we first exported a list of differential expressions related to this gene. We found 553 up-regulated and 345 down-regulated genes that were significantly

associated with STAT5B expression ( $P_{adj} < 0.05$ ,  $|\log_2(FC)| > 1.5$ , Fig. 5A). Then, enrichment analysis of GO and KEGG pathways on differentially expressed genes related to STAT5B was performed, and we obtained 254 biological processes, 44 cellular components and 5 molecular functions ( $P_{adj} < 0.05$ ,  $qvalue < 0.2$ ). The functions associated with AML are illustrated in Figs. 5B.

## Correlation between STAT5B expression and immune infiltration in AML

The single sample Gene Set Enrichment Analysis (ssGSEA) algorithm in R software is used to assess the correlation between STAT5B and immune cells in leukemia tissues. The analysis results are presented in Fig. 6. STAT5B expression was positively correlated with infiltration of T helper cells, Mast cells, T cells, Tcm, CD8 T cells, Cytotoxic cells, NK cells, and B cells, while it was negatively correlated with infiltration of Dendritic cells, Macrophages, and Eosinophils. These studies indicate that the STAT5B expression is associated with immune infiltration in AML.

## Discussion

The family of STAT genes is thought to originate from successive genome duplication and functional divergence of a single ancestral STAT gene early in vertebrate evolution. Typically, each of the STAT family members is a target gene induced when the respective transcription factor is activated. Many upstream kinases can activate STATs, though the most common is the Janus kinase, which associate with various cell surface receptors. Janus kinases are a family of non-receptor tyrosine kinases that primarily transduce signals from the cell surface to the nucleus by activating downstream STATs. Subsequently, STATs translocate to the nucleus and bind with DNA to transcribe genes that are involved in cellular proliferation, survival, angiogenesis and immunity. In cancer, STATs can become activated constitutively due to aberrant activation of an upstream kinase and/or loss of a negative regulator, leading to the changes in gene expression that underlies the malignant phenotype. Although the role of STATs in the occurrence and prognosis of certain cancers has been clarified, there are very few reports on the expression and role of different STATs in AML. In this study, we explored the transcription level, molecular mechanism, and biological function of STATs in AML and its relationship with prognosis and immune infiltration in patients with AML.

Based on the available evidence, the role of STAT1 in patients with cancer remains controversial. Moreover, the mechanism behind the oncogenic or tumor-suppressing role of STAT1 remains unclear. STAT1 deficiency studies show that STAT1 act as a tumor promoter in leukemia development(25). Of the STAT family, only STAT1 and STAT2 are critical for the IFN pathway(26). However, little is known about the function of STAT2 in malignancy. Similar to STAT2, previous studies on STAT4 in malignancy are limited. Moreover, functions of STAT4 in carcinogenesis and tumor metastasis are paradoxical. STAT6 is a key member of the STAT family, whose role in the biology of lymphoma and leukemia has been firmly established(27). Our research shows that STAT1, STAT2, STAT4 and STAT6 are highly expressed in AML, but only STAT6 has an effect on survival. Nevertheless, the expression levels of STAT3, STAT5A, and

STAT5B were lower in AML than in normal tissues. Although STAT3 activation is transient and dispensable in normal cells, it often occurs constitutively in cells of many hematological (and solid) cancers, which depend on STAT3 for survival(28). STAT3 is constitutively activated in various human cancers, and its activation is frequently associated with poor prognosis. STAT5 has two highly homologous isoforms, STAT5A and STAT5B. Although STAT5 has been implicated in several cancers for tumor proliferation, apoptosis and invasion, the majority of studies reported STAT5 genes predicted a favorable clinical outcome in malignancy. In this study, we found that STAT3 levels have no bearing on the prognosis of AML patients. In addition, our current results indicate that the high expression of STAT5B is related to the improvement of OS in AML patients, but there is no correlation between the expression of STAT5A and OS in AML patients. Liu et al. reported that expression levels of STAT2, STAT4 and STAT6 were significantly higher in AML than in normal tissues. Furthermore, high expression of STAT6 was correlated to lower OS in AML patients(29). Intriguingly, our study shows that more members of STATs are differentially expressed between AML and normal tissues. In addition, we found that high expression of STAT6 is associated with lower OS, and high expression of STAT5B is associated with better OS. This can be attributed to different study design, sample size and cutoff. However, by multivariate analyses and constructing the LASSO model, we proved that STAT5B expression is an independent marker for the prognosis of AML. To the best of our knowledge, this is the first report investigating the relationship between STAT5B and prognosis of AML.

We assessed the function of STATs by GO enrichment analysis and KEGG pathway enrichment in human. Then we used DisGeNet and found that STATs are highly correlated with myeloid leukemia. DisGeNET is a newly developed platform that covers a comprehensive catalog of genes and variants related to human diseases(30). It collected 4145 genes related to 738 tumor diseases, deepening our understanding of the link between genetic variation and cancer incidence and progression. Through the single-gene difference combined enrichment analysis, we predicted functions and pathways that STAT5B may participate in. Studies have revealed that the function of STAT5B is significantly related to the occurrence, development and tumor microenvironment of AML.

In solid tumors, the tumor genetic environment directly regulates the anti-tumor immune response and the efficacy of immunotherapy(31). The genetic changes of AML cells have been extensively studied and deeply understood, but there are few studies on the immunological characteristics of AML. The result of this study is that the expression level of STAT5B was positively correlated with infiltration of T cells, while it was negatively correlated with infiltration of dendritic cells. The key step in activating the immune system is to recognize the surface antigens of tumor cells. Because T cells naturally traffic to hematopoietic sites where AML localizes, T-cell therapy has significant potential. AML cells are essentially the precursors of dendritic cells antigen presenting cells, but their ability to produce effective immune responses is weak. This may be explained by the inability to form effective immune synapses and present antigens to T cells(32). Comprehensive immune profiling of newly diagnosed AML patients proved that the functional defects of T cells and NK cells are the main aspects of immune dysfunction, while the function of B cells is not affected(33). The senescence and exhaustion of T cells, the over-

maturation of NK cells and impaired function of  $\gamma\delta$ T cells all participate in immunosuppression, causing AML cells to evade immune attack.

## Conclusions

In summary, our research demonstrates that the transcription level of STATs varies in AML. In addition, abnormal expressions of STAT5B and STAT6 were considered to be related with clinical outcomes in patients with AML. The study found that STAT5B can be used as potential survival prognostic biomarkers and targets for AML. However, more experimental studies are needed to confirm our results, thereby promoting the clinical application of STAT5B as a prognostic indicator or immunotherapy target in AML.

## Methods

### Metascape analysis

Metascape (<https://metascape.org/>) is a web-based portal designed to deliver a comprehensive gene list annotation and analysis resource for experimental biologists. In terms of design features, Metascape combines functional enrichment, interaction analysis, gene annotation, and membership search to leverage over 40 independent knowledge bases within one integrated portal(34). The functions of STATs were predicted by analyzing GO and KEGG in Metascape. Then, we used Metascape to perform enrichment analysis on STATs and identified their relationship with human diseases in DisGeNET(35). Only terms with P-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 were considered as significant.

### TCGA database analysis

To investigate the expression pattern of STATs in AML, UCSC XENA (<https://xenabrowser.net/datapages/>) was utilized. RNA-sequencing data in transcripts per million reads (TPM) format of TCGA and Genotype-Tissue Expression (GTEx) processed uniformly with the Toil process(36). Extract the corresponding normal tissue data in TCGA's LAML and GTEx. RNA-sequencing data in TPM format and log<sub>2</sub> conversion for STATs expression comparison between in AML and normal samples. The above analysis uses of R software package ggplot2.

Raw counts of RNA-sequencing data (level 3) and corresponding clinical information from 151 AML samples were obtained from TCGA dataset (<https://portal.gdc.cancer.gov/>) in January 2020, in which the method of acquisition and application complied with the guidelines and policies. According to the median value of STATs expression, patient samples were divided into high expression group and low expression group. The prognostic effect of STATs expression was analyzed through Kaplan-Meier analysis using the Log rank test. Univariate and multivariate proportional hazard regression analysis was carried out using Cox regression. The prognostic value of STATs expression in AML patients and their OS were assessed by R software package survival, and then visualized by R software package survminer.

The Kaplan-Meier survival analysis with log-rank test was also used to compare the survival difference between different STATs. TimeROC analysis was performed to compare the predictive accuracy of each gene and risk score. The LASSO regression algorithm for feature selection, using 10-fold cross-validation. The above analysis uses the R software package glmnet. For Kaplan–Meier curves, p-values and the hazard ratio (HR) with 95% confidence interval (CI) were generated by log-rank tests and univariate Cox proportional hazards regression.  $P < 0.05$  was considered as statistically significant.

## **Functional Analysis**

After deducing the most prognostic gene in the STAT family, we divide it into high and low expression groups and analyze the differential expression of the two groups, in order to get a list of differential expressions linked to this gene. Then we performed GO and KEGG pathway enrichment analysis to further infer the function or pathway that the gene may be associated with. The single gene expression difference analysis uses the R software package DESeq2(37). The enrichment analysis was assessed by R software package clusterProfiler(38), and then visualized by R software package ggplot2.

## **Immune infiltration analysis**

Level 3 RNA-sequencing data in fragments per kilobase million reads (FPKM) format and corresponding clinical information of TCGA AML project were downloaded. RNA-sequencing data in FPKM format are analyzed after log<sub>2</sub> conversion. The ssGSEA algorithm is used to calculate the score of immune cells in the tissue and infer the infiltration of immune cells in the tissue. The ssGSEA algorithm is incorporated into the R software package GSVA(39). 24 immune cell markers are from an Immunity article(40). All analytical methods above and R packages were performed using R software version 3.6.3.

## **Abbreviations**

<i>STAT</i>	<i>signal transducer and activator of transcription</i>
AML	acute myeloid leukemia
OS	overall survival
LASSO	least absolute shrinkage and selection operator
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
TCGA	The Cancer Genome Atlas
ssGSEA	single sample Gene Set Enrichment Analysis
TPM	transcripts per million reads
GTE <sub>x</sub>	Genotype-Tissue Expression
HR	hazard ratio
CI	confidence interval
FPKM	fragments per kilobase million reads

## Declarations

### Ethics approval and consent to participate

Ethics committee approval was issued by the Institutional Ethics Committee of Huaihe Hospital of Henan University to the commencement of the study.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

### Funding

There is not any funding to report.

### Authors' contributions

TQ conceived the study, participated in its design and coordination, and collected data. YS mainly participated in statistical analysis, and drafted the manuscript. HZ contributed to the conception of the work and the interpretation of study data. All authors have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

## Acknowledgments

The author would like to acknowledge the TCGA project for data sharing.

## References

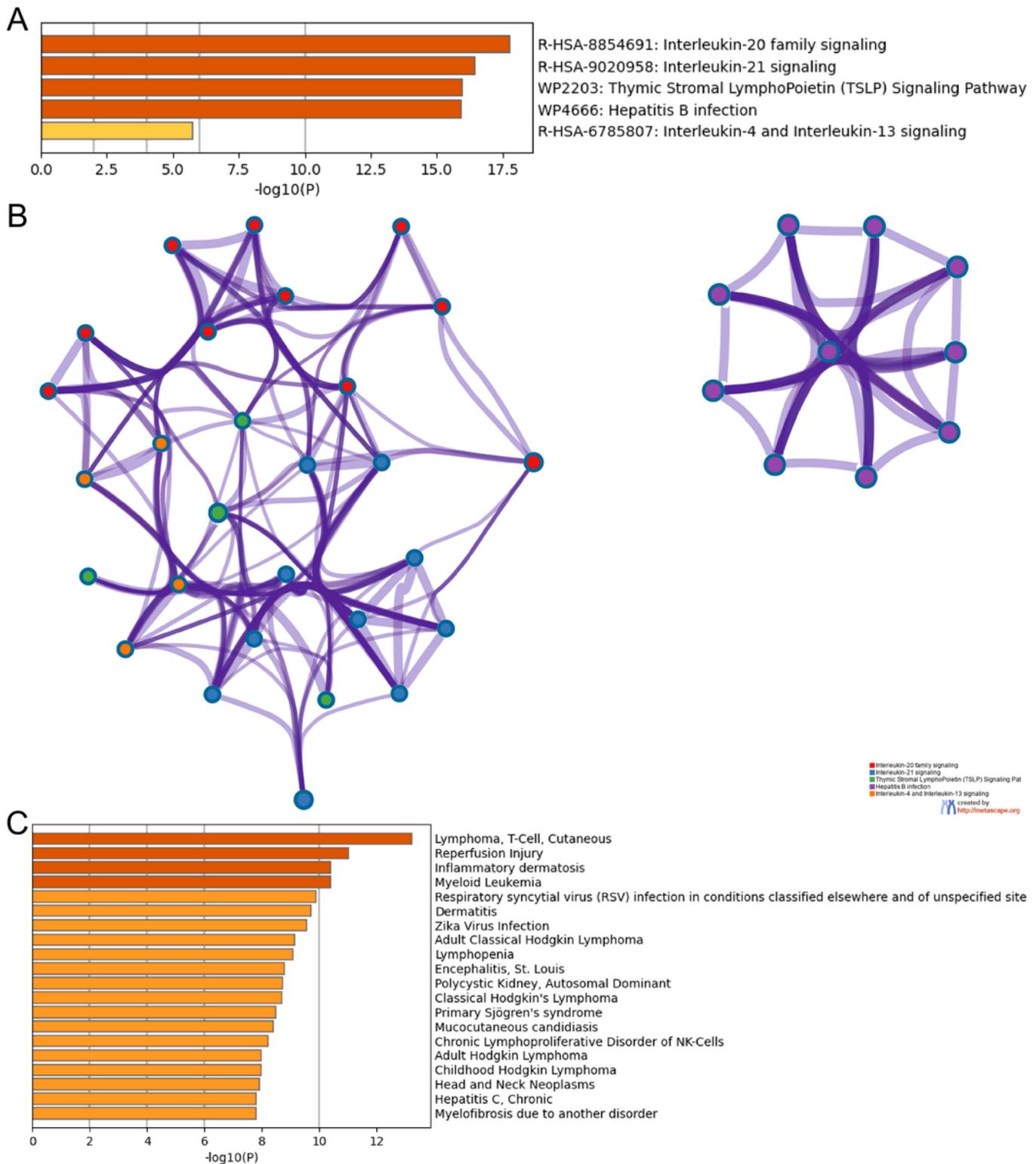
1. Dohner H, Estey EH, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-47.
2. Bullinger L, Döhner K, Döhner H. Genomics of Acute Myeloid Leukemia Diagnosis and Pathways. *J Clin Oncol*. 2017;35(9):934-46.
3. Arber DA, Orazi A, Hasserjian RP, Thiele J, Borowitz MJ, Beau MML, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-405.
4. Green CL, Koo K, Hills RK, Burnett AK, Linch DC, Gale RE. Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. *J Clin Oncol*. 2010;28(16):2739-47.
5. Schlenk RF, Kayser S, Bullinger L, Kobbe G, Casper J, Ringhoffer M, et al. Differential impact of allelic ratio and insertion site in FLT3-ITD-positive AML with respect to allogeneic transplantation. *Blood*. 2014;124(23):3441-9.
6. Stengel A, Kern W, Haferlach T, Meggendorfer M, Fasan A, Haferlach C. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia*. 2016.
7. Gaidzik VI, Teleanu, Papaemmanuil E, Weber D, Paschka P, Hahn J, et al. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*. 2016;30(11):2160-8.
8. Paschka P, Schlenk RF, Gaidzik VI, Herzig JK, Aulitzky T, Bullinger L, et al. ASXL1 mutations in younger adult patients with acute myeloid leukemia: a study by the German-Austrian Acute Myeloid Leukemia Study Group. *Haematologica*. 2015;100(3):324-30.

9. Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S, et al. Simpson's Paradox and the Impact of Different DNMT3A Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia. *J Clin Oncol*. 2015;33(18):2072-83.
10. Loh C-Y, Arya A, Naema AF, Wong WF, Sethi G, Looi CY. Signal Transducer and Activator of Transcription (STATs) Proteins in Cancer and Inflammation: Functions and Therapeutic Implication. *Frontiers in oncology*. 2019;9:48-.
11. Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A Novel Transcription Factor, Tbet, Directs Th1 Lineage Commitment. *Cell*. 2000;100(6):655-69.
12. Ma H, Lu C, Ziegler J, Liu A, Sepulveda A, Okada H, et al. Absence of Stat1 in donor CD4+ T cells promotes the expansion of Tregs and reduces graft-versus-host disease in mice. *J Clin Invest*. 2011;121(7):2554-69.
13. O'Shea JJ, Holland SM, Staudt LM. JAKs and STATs in immunity, immunodeficiency, and cancer. *N Engl J Med*. 2013;368(2):161-70.
14. Zhang Y, Chen Y, Liu Z, Lai R. ERK is a negative feedback regulator for IFN- $\gamma$ /STAT1 signaling by promoting STAT1 ubiquitination. *BMC Cancer*. 2018;18(1):613.
15. Chen Y, Fang L, Zhang J, Li G, Ma M, Li C, et al. Blockage of Glyoxalase I Inhibits Colorectal Tumorigenesis and Tumor Growth via Upregulation of STAT1, p53, and Bax and Downregulation of c-Myc and Bcl-2. *Int J Mol Sci*. 2017;18(3):570.
16. Zhang J, Wang F, Liu F, Xu G. Predicting STAT1 as a prognostic marker in patients with solid cancer. *Ther Adv Med Oncol*. 2020;12:1758835920917558-.
17. Chowdhury FZ, Farrar JD. STAT2: A shape-shifting anti-viral super STAT. *JAKSTAT*. 2013;2(1):e23633-e.
18. Bai L, Zhou H, Xu R, Zhao Y, Chinnaswamy K, McEachern D, et al. A Potent and Selective Small-Molecule Degradator of STAT3 Achieves Complete Tumor Regression In Vivo. *Cancer Cell*. 2019;36(5):498-511.e17.
19. Busker S, Qian W, Haraldsson M, Espinosa B, Johansson L, Attarha S, et al. Irreversible TrxR1 inhibitors block STAT3 activity and induce cancer cell death. *Sci Adv*. 2020;6(12):eaax7945.
20. Xiang M, Kim H, Ho VT, Walker SR, Bar-Natan M, Anahtar M, et al. Gene expression-based discovery of atovaquone as a STAT3 inhibitor and anticancer agent. *Blood*. 2016;128(14):1845-53.
21. Yang C, Mai H, Peng J, Zhou B, Hou J, Jiang D. STAT4: an immunoregulator contributing to diverse human diseases. *Int J Biol Sci*. 2020;16(9):1575-85.

22. Ferbeyre G, Moriggl R. The role of Stat5 transcription factors as tumor suppressors or oncogenes. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2011;1815(1):104-14.
23. Maier E, Duschl A, Horejs-Hoeck J. STAT6-dependent and -independent mechanisms in Th2 polarization. *Eur J Immunol*. 2012;42(11):2827-33.
24. Karagiannis TT, Cleary JP, Gok B, Henderson AJ, Martin NG, Yajima M, et al. Single cell transcriptomics reveals opioid usage evokes widespread suppression of antiviral gene program. *Nature Communications*. 2020;11(1):2611.
25. Kovacic B, Stoiber D, Moriggl R, Weisz E, Ott RG, Kreibich R, et al. STAT1 acts as a tumor promoter for leukemia development. *Cancer Cell*. 2006;10(1):77-87.
26. Ashour J, Morrison J, Laurent-Rolle M, Belicha-Villanueva A, Plumlee CR, Bernal-Rubio D, et al. Mouse STAT2 Restricts Early Dengue Virus Replication. *Cell Host Microbe*. 2010;8(5):410-21.
27. Bruns HA, Kaplan MH. The role of constitutively active Stat6 in leukemia and lymphoma. *Critical Reviews in Oncology/Hematology*. 2006;57(3):245-53.
28. Frank DA. STAT3 as a central mediator of neoplastic cellular transformation. *Cancer Letters*. 2007;251(2):199-210.
29. Liu W, Zhu F, Yan J, Liu Y, Chen C, Zhang K, et al. Identification and Validation of STAT6 as a Prognostic and Predictive Biomarker in Acute Myeloid Leukemia. *Onco Targets Ther*. 2020;13:11165-76.
30. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*. 2017;45(D1):D833-D9.
31. Chowell D, Krishna C, Pierini F, Makarov V, Rizvi NA, Kuo F, et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat Med*. 2019;25(11):1715-20.
32. Geiger TL, Rubnitz JE. New approaches for the immunotherapy of acute myeloid leukemia. *Discov Med*. 2015;19(105):275-84.
33. Williams P, Basu S, Garcia-Manero G, Hourigan CS, Oetjen KA, Cortes JE, et al. The distribution of T-cell subsets and the expression of immune checkpoint receptors and ligands in patients with newly diagnosed and relapsed acute myeloid leukemia. *Cancer*. 2019;125(9):1470-81.
34. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*. 2019;10(1):1523.

35. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*. 2016;45(D1):D833-D9.
36. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nature Biotechnology*. 2017;35(4):314-6.
37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550.
38. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-7.
39. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7.
40. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf Anna C, et al. Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity*. 2013;39(4):782-95.

## Figures



**Figure 1**

Enrichment analysis of STATs in human sapiens (Metascape). (A) Bar graph of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enriched terms, colored by P-values. (B) Network of GO and KEGG enriched terms color by cluster ID, where nodes that share the same cluster ID are typically close to each other. (C) Summary of enrichment analysis in DisGeNET.

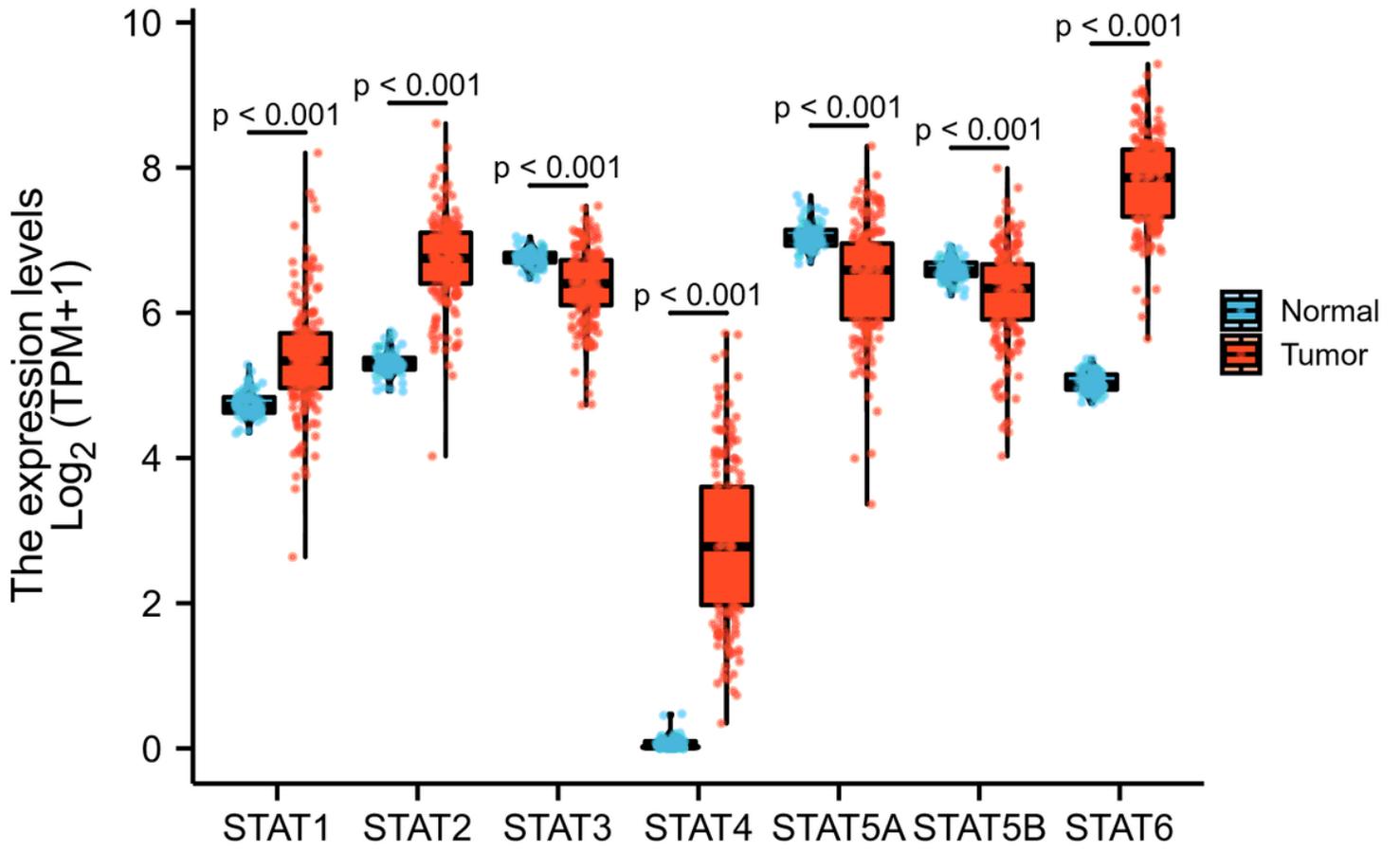
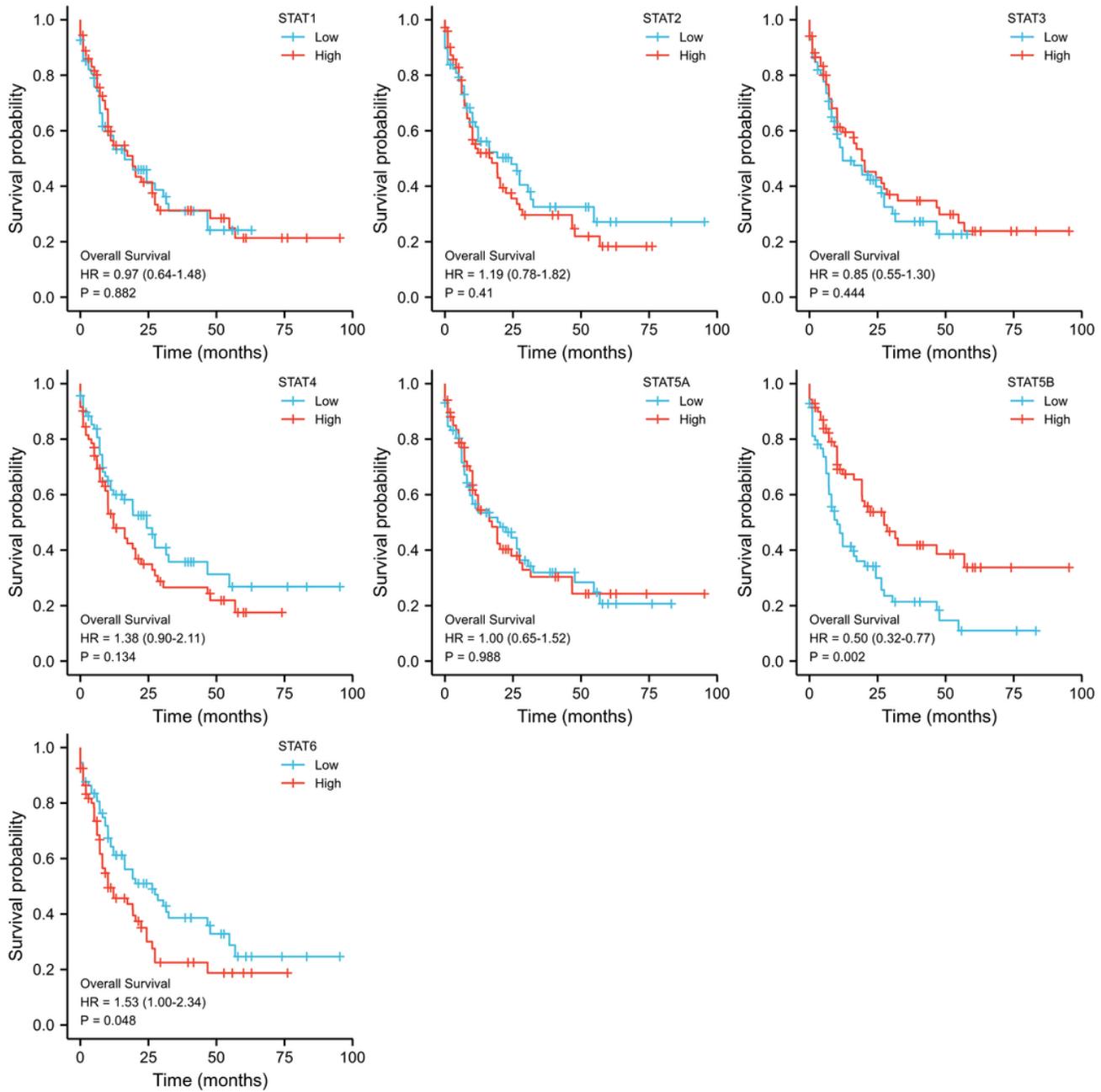


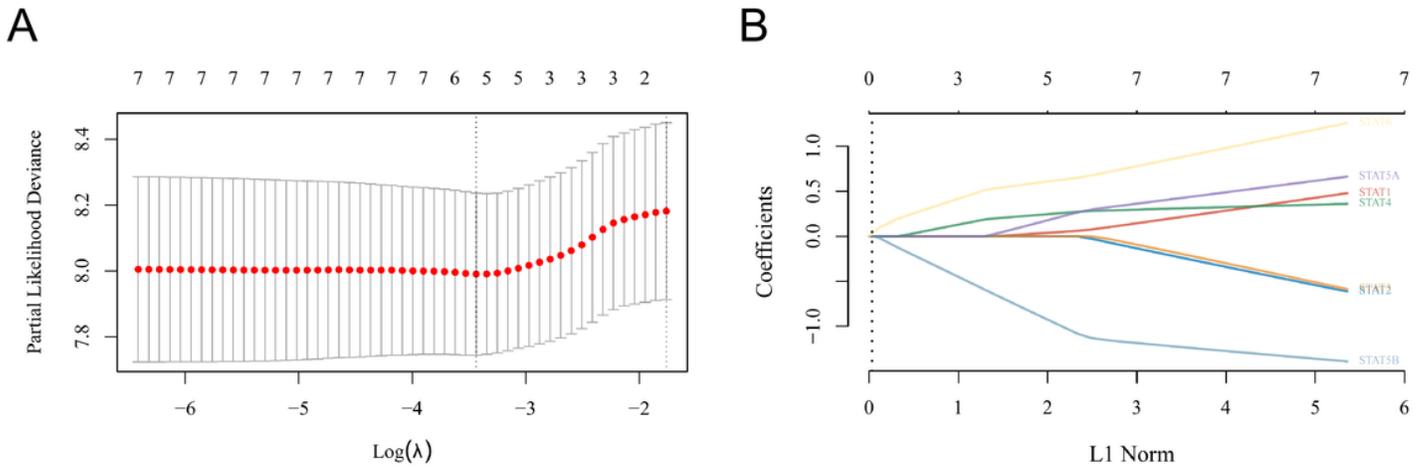
Figure 2

The expression of STATs in AML.



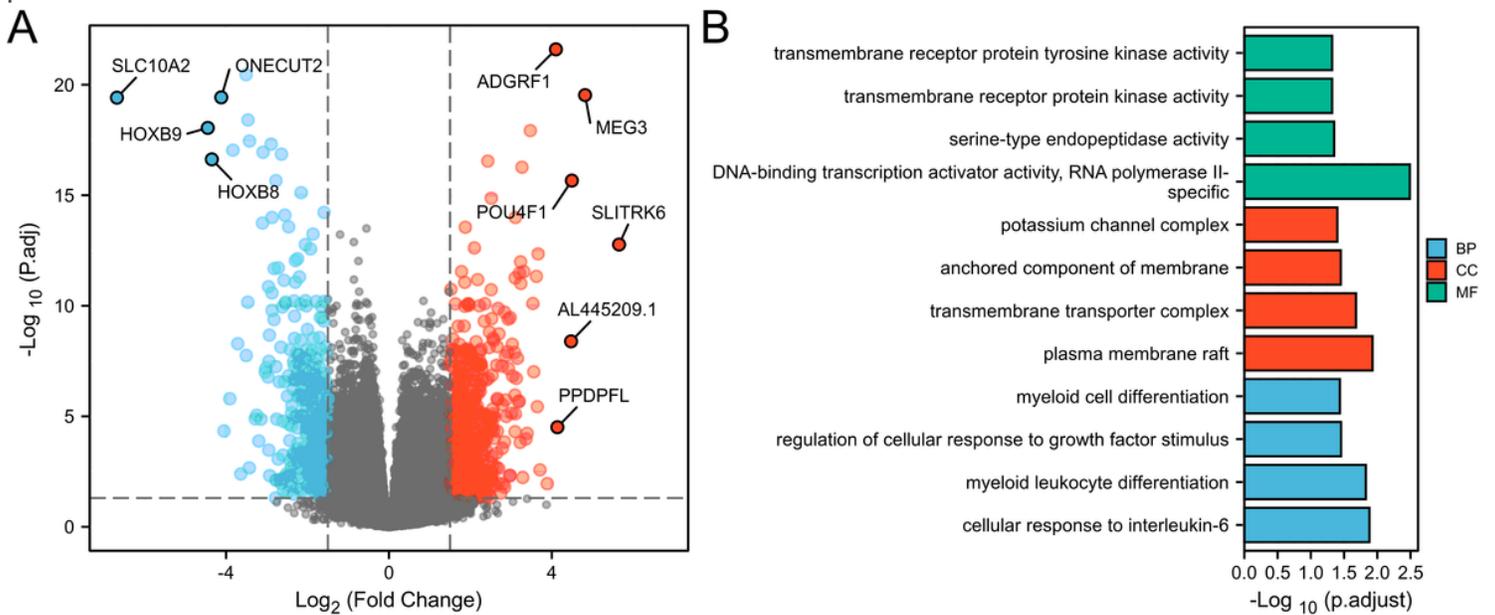
**Figure 3**

OS curve of different expressed STATs in AML patients.



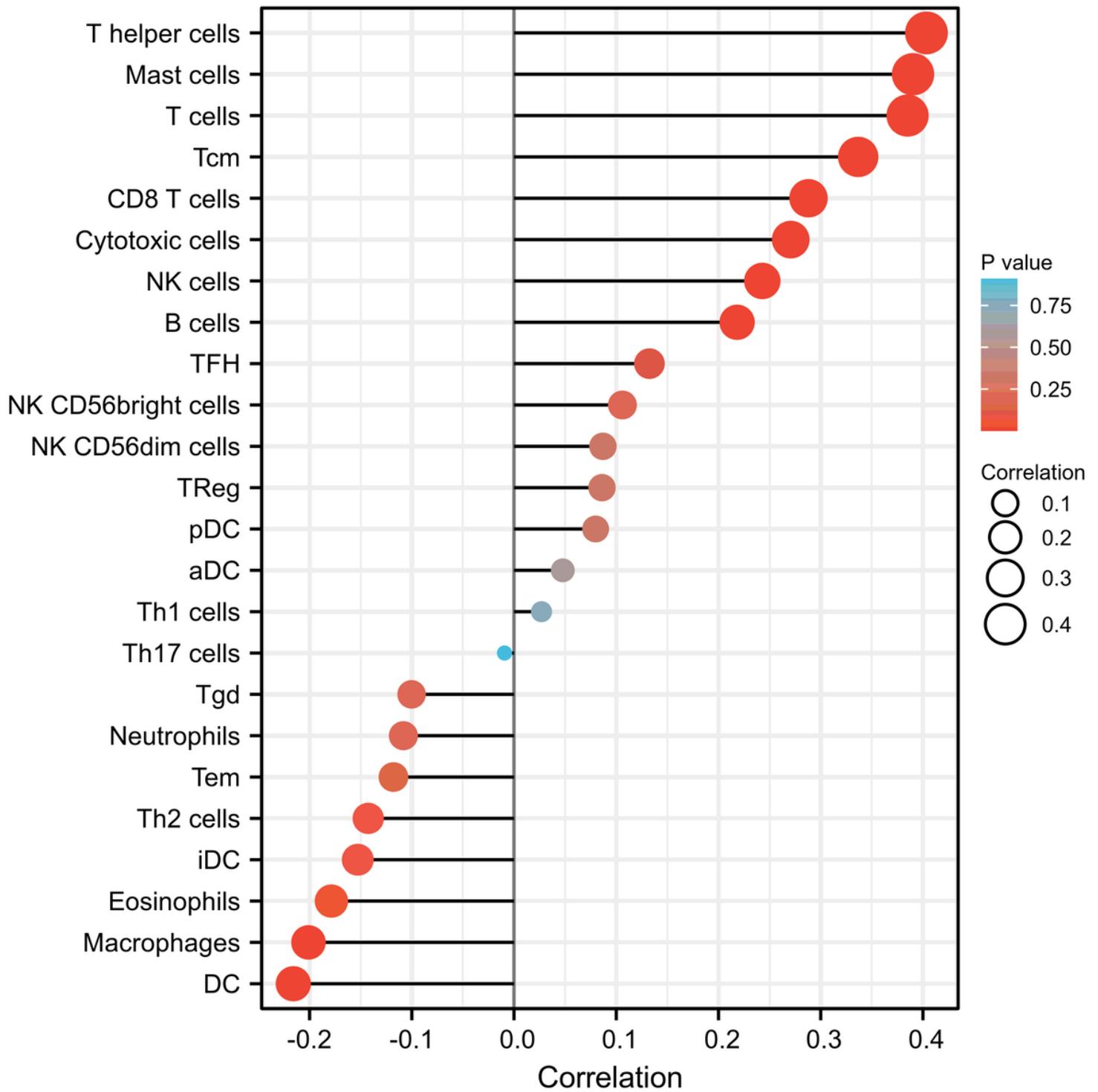
**Figure 4**

Survival analysis in AML prognosis prediction. (A) Partial likelihood deviance versus  $\text{log}(\lambda)$  was drawn using LASSO Cox regression model. (B) Coefficients of selected features are shown by lambda parameter.



**Figure 5**

Functions and signal pathways related to the differential expression of STAT5B in AML. (A) Volcano plot of differential gene expression related to STAT5B. Up-regulation and down-regulation were marked by red and blue circles, respectively. (B) Enrichment analysis of differential gene related to STAT5B in AML.



**Figure 6**

The relationship between STAT5B expression levels and immune infiltration levels in AML.